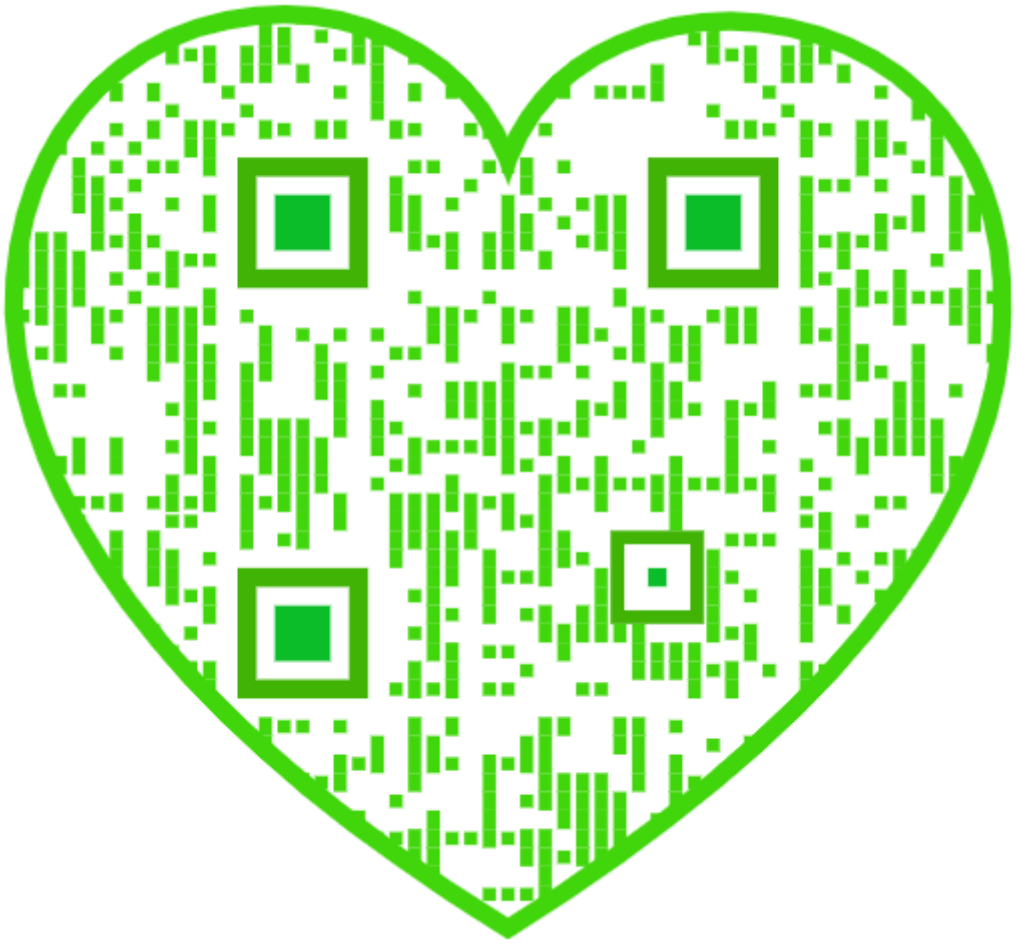
Master in Artificial Intelligence

Feature Engineering II



Purpose

The purpose of the section is to help you learn how to identify and extract meaningful features from the data to become a Successful Artificial Intelligence (AI) Engineer

At the end of this lecture, you will learn the following

How to engineer new features or transform existing features

How to engineer new features or transform existing features

Understand the Problem and Data

Perform Exploratory Data Analysis (EDA)

Feature Engineering

Feature Selection

Dimensionality Reduction

Evaluate Feature Importance

Iterate and Experiment

Domain Knowledge Integration

Validate and Test

One-Hot Encoding

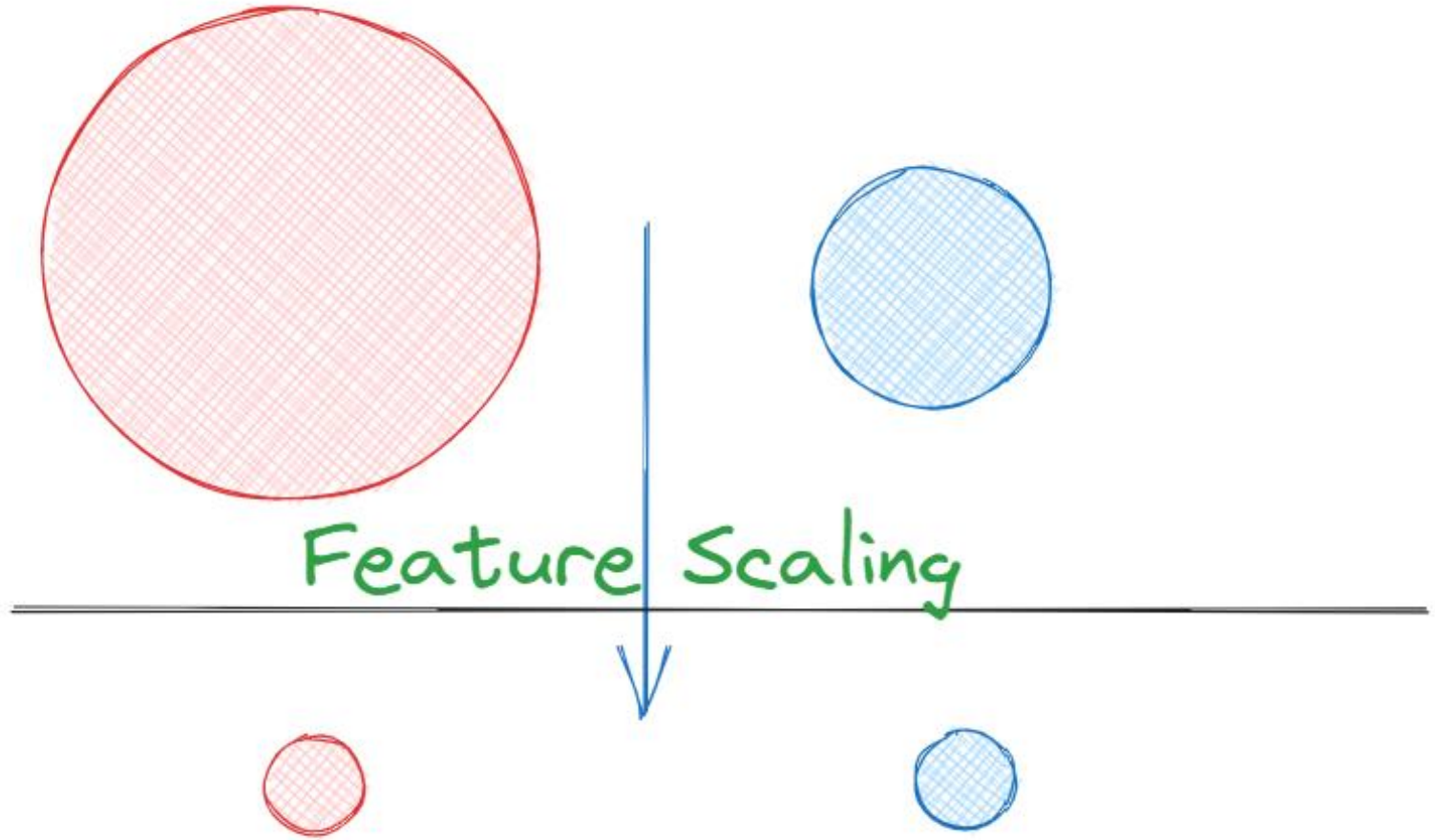
Original

Gender
Male
Female
Male
Male

One-hot encoded

Gender	Male	Female
Male	1	0
Female	0	1
Male	1	0
Male	1	0

Feature Scaling



Feature Normalization

Mean normalization

Replace x_i with $x_i - \mu_i$ to make features have approximately zero mean (Do not apply to $x_0 = 1$).

E.g. $x_1 = \frac{size - 1000}{2000}$

Average size = 100

$x_2 = \frac{\#bedrooms - 2}{5}$ 4

1 - 5 bedrooms

$-0.5 \leq x_1 \leq 0.5$, $-0.5 \leq x_2 \leq 0.5$

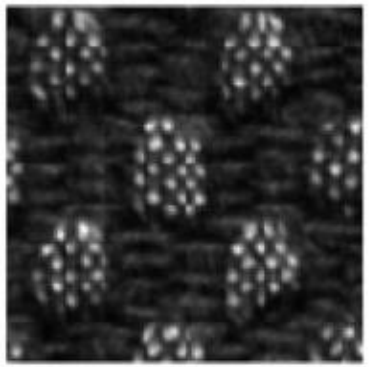
$x_1 \leftarrow \frac{x_1 - \mu_1}{s_1}$ (μ_1: avg value of x_1 in training set; s_1: range (max - min) (or standard deviation))

$x_2 \leftarrow \frac{x_2 - \mu_2}{s_2}$

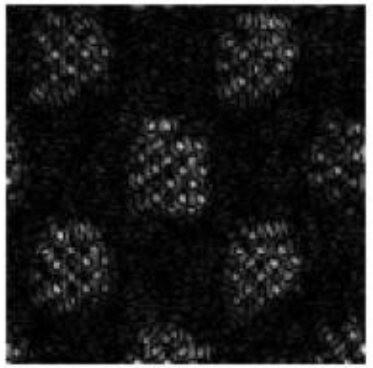
Andrew Ng

Feature Aggregation

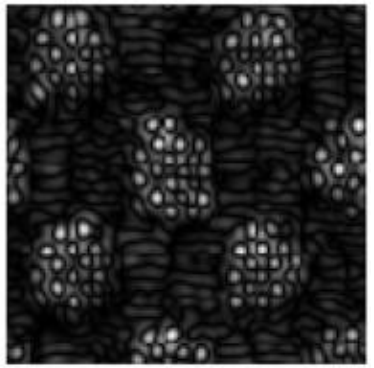
initial image



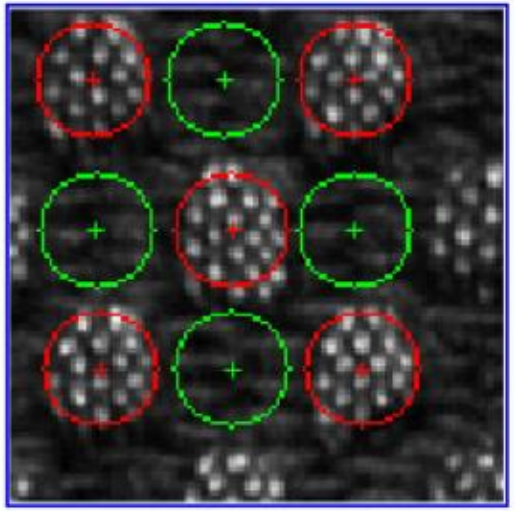
feature map 1



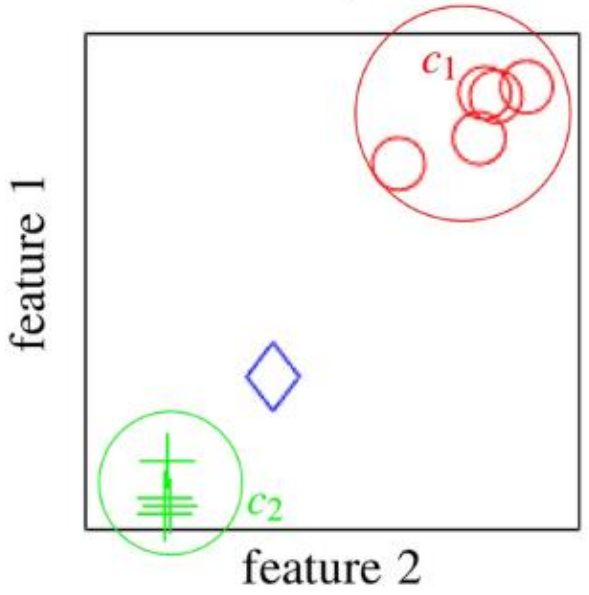
feature map 2



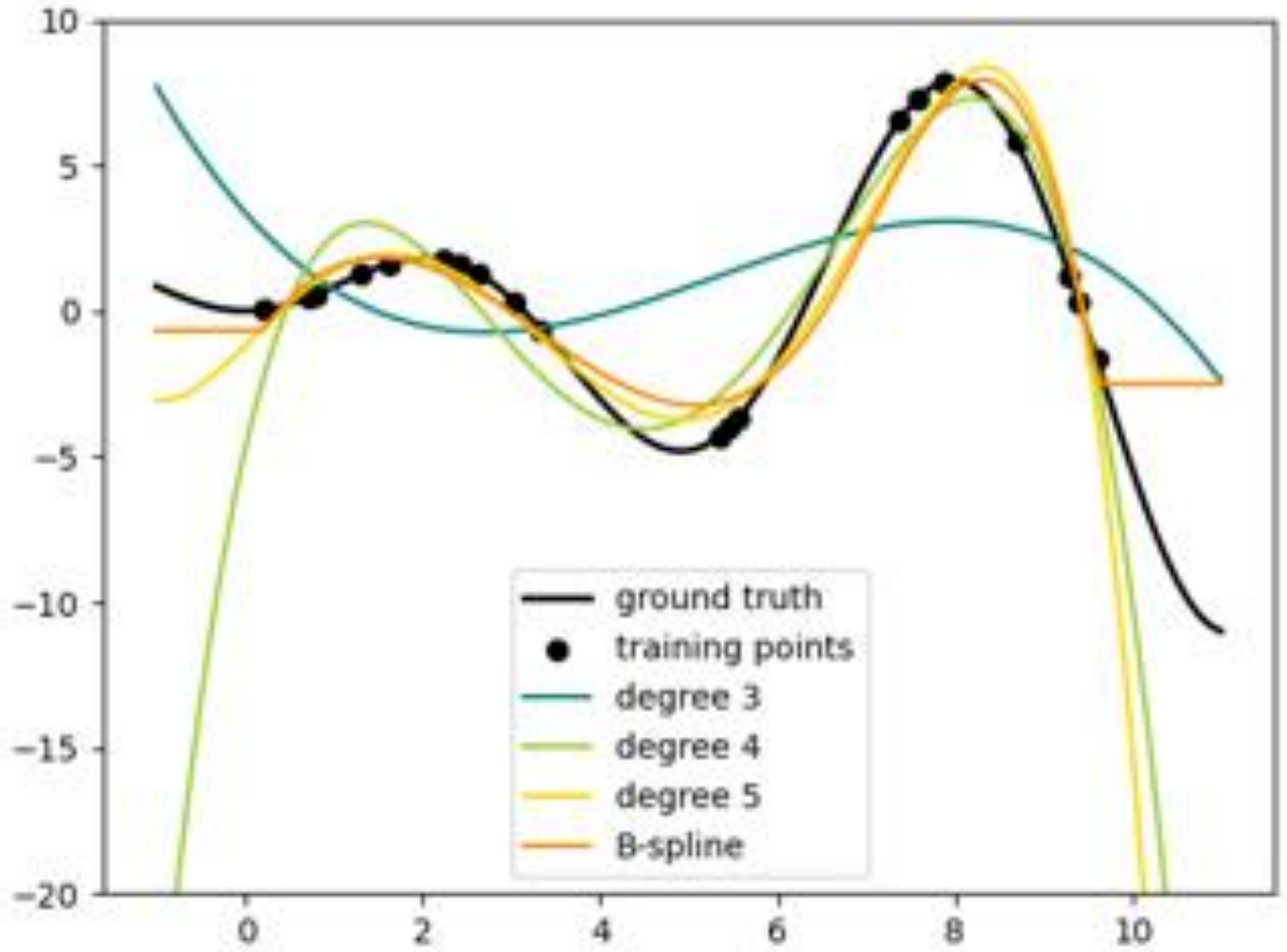
regions M_1, M_2, M_3



feature space



Polynomial Features



Polynomial Features

- Import Libraries
- Prepare Data
- Instantiate PolynomialFeatures
- Fit and Transform
- Convert to DataFrame
- Review Results
- Use in Modeling

Feature Interaction

- Identify Relevant Features
- Feature Combination
- Feature Interaction
- Feature Engineering
- Evaluate and Validate
- Iterate and Experiment

Domain-specific Features

Understand the Problem Domain

↓

Collaborate with Domain Experts

↓

Identify Relevant Variables

↓

Feature Extraction

↓

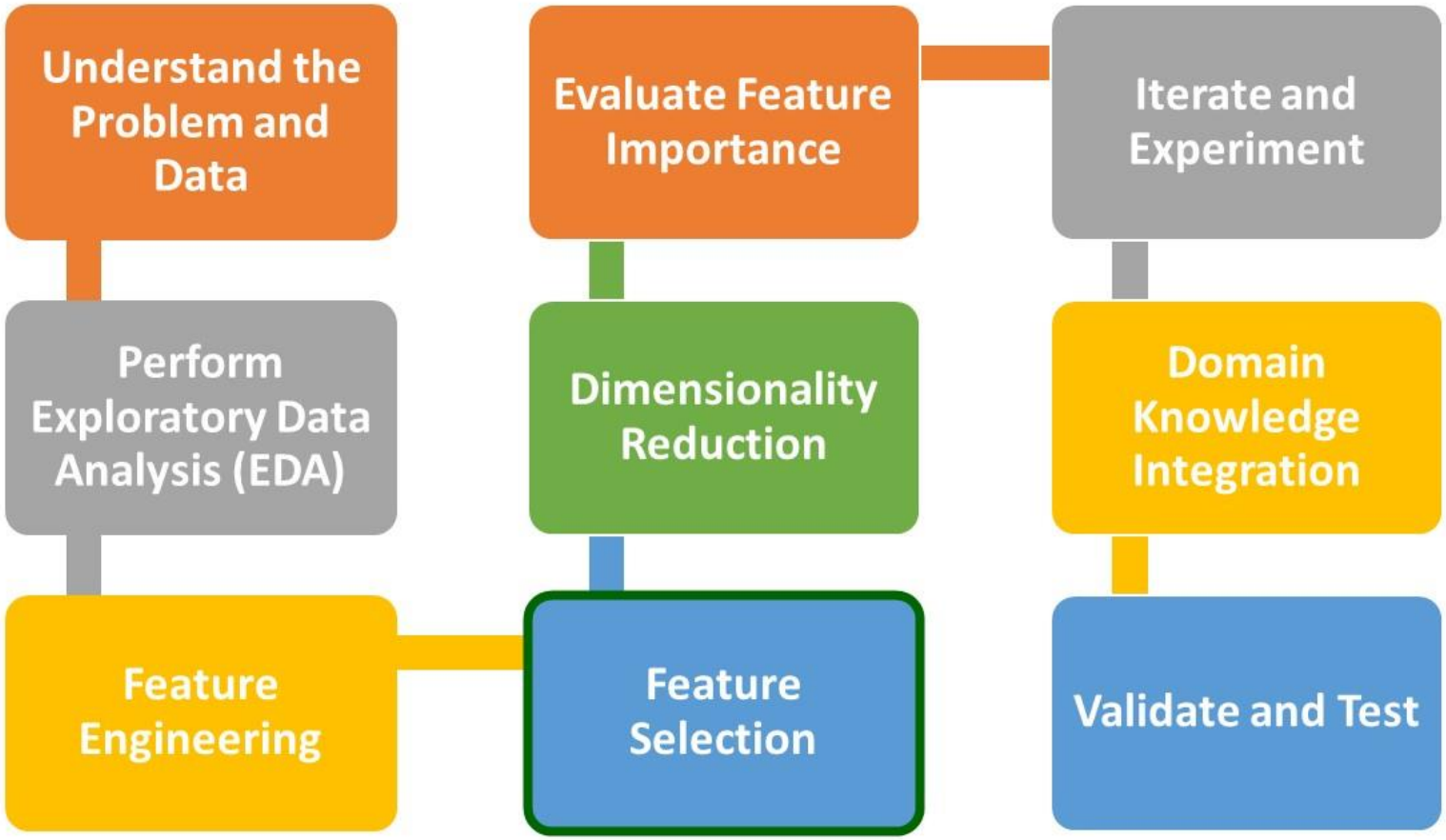
Feature Encoding

Domain-specific Features

- Temporal Features
- Geospatial Features
- Domain-Specific Transformations
- Iterate and Validate
- Collaborate and Communicate

What is next?

How to select a subset of the most relevant features



Master in Artificial Intelligence

Feature Engineering II